2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)
July 6-11, 2014, Beijing, China

Aeroengine prognosis through Genetic Distal Learning applied to uncertain Engine Health Monitoring data

Alvaro Martínez
Rolls-Royce Deutschland Ltd & Co KG
Blankenfelde-Mahlow, Germany
Email: Alvaro.Martinez@Rolls-Royce.com

Luciano Sánchez
Universidad de Oviedo
Gijón, Asturias, Spain
Email: luciano@uniovi.es

Inés Couso
Universidad de Oviedo
Gijón, Asturias, Spain
Email: couso@uniovi.es

***Abstract*—Genetic Fuzzy Systems have been successfully applied to assess Engine Health Monitoring (EHM) data from aeroengines, not only due to their robustness towards noisy gas path measurements and engine-to-engine variability, but also because of their capability to produce human-readable expressions. These techniques can detect the presence of certain types of abnormal events or specific engine conditions, where a combination of the EHM signals only appears when these occur. However, an engine that repeatedly operates under unfavourable conditions will also have a reduced life. Smooth deteriorations do no manifest themselves as combinations of the EHM signals, the current existing techniques can therefore not assess these. In this paper it is proposed to use distal learning to build a model that indirectly identifies the deterioration rate of an aeroengine. It will be shown that the integral of the modelled rate is a prognostic indicator of the remaining life of the engine to a selected end condition. The results are subsequently tested on a representative sample of aeroengine data.**

Keywords: Engine Health Monitoring; Genetic Fuzzy Systems; Distal Learning

I. Introduction

Equipment Health Monitoring (EHM) is the assessment of engine instrumentation data over time in order to detect substantial anomalies or incipient events. The application of prognostics within an EHM management system are intended to estimate the remaining life of an engine, anticipating certain events or findings and therefore reducing the number or degree of engine refurbishments [6]. The assessment of EHM data not only reviews the individual working conditions but also the trend over time in order to identify rapid levels of deterioration. Often, a comparison is made of the engine data against those parameters identified to be characteristic of known engine conditions or against design limits [15]. However understanding the design limits for a new engine or predicting the engine parameter deterioration levels over time is complex and several methods have been developed.

A. EHM assessment existing models

The most common EHM assessment methods are based around Gas Path Analysis (GPA). The gas path components are all air-washed parts within the engine gas path, the compressors, the combustor and the turbines (see Figure 1). The gas path components are susceptible to distinct different issues, such as worn seals, excessive tip clearances, burning, cracking or missing parts or sections of parts, etc. (see Figure 2). The purpose of GPA is to detect changes in the internal working conditions of the engine as early as possible through the observation of EHM parameters [15]. Standard assessment methods which review the engine development over time include deterioration modelling and probabilistic simulation [9]. Recently, assessments have made use of fuzzy logic and neural networks to develop new pattern recognition methods to identify engine trends and step changes [5][7][8].

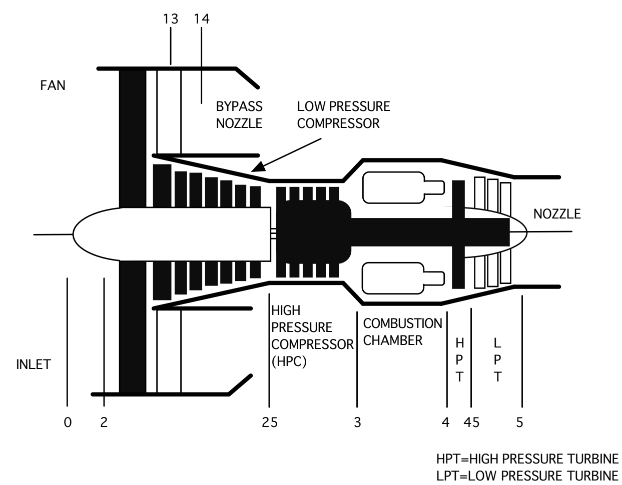

Fig. 1. Typical two shaft high bypass ratio turbo fan.

The main objective of this type of assessments is to determine the optimum engine maintenance interval and assure appropriate levels of reliability for the fleet. The introduction of maintenance contracts as Power-By-The-Hour where the management of the engine maintenance is the responsibility of the OEMs, has emphasized the need for the early diagnosis of engine specific deterioration. This is, further development in the assessment of EHM data has been highlighted so that small shifts and trends in the variables are identified, even when the values are still within the appropriate reliability levels of the specific parameter. This way, the level of engine deterioration at the time of the engine maintenance may be determined in advance and the prioritization within the fleet performed ahead of time based not just on average fleet experience but on also on each engines' own specific level of deterioration.

978-1-4799-2072-3/14/$31.00 ©2014 IEEE

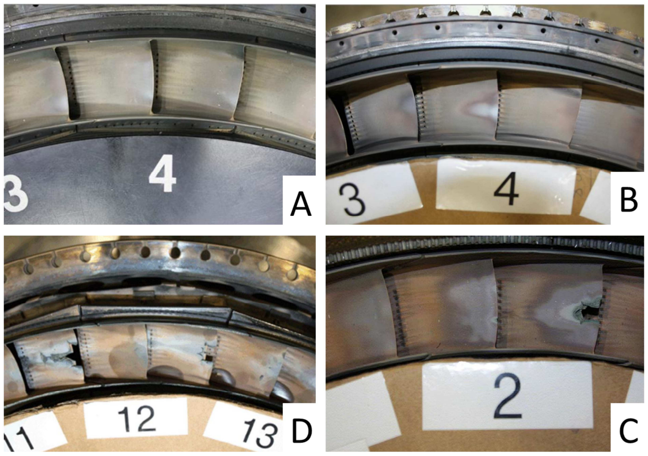

Fig. 2. General deterioration over time of a turbine nozzle guide vane. Clockwise it can be seen how the vane is still deemed to be in a serviceable condition (A), it evolves to a repairable condition (B), however it is then deemed to be scrap (C) and ultimately it is considered to directly affect the engine working condition (D)..

B. Uncertainty in EHM data

Flight conditions as well as the internal condition of each individual engine influence gas path measurements. In order to reduce some of the variability between engines, EHM data is typically not expressed in absolute values. The managed EHM data from an engine is estimated from the deltas between the engine's own measurements and those from a known, baseline engine. In addition, different techniques are available, which have been used to filter out the noise in the EHM data [14].

Engine events or significant engine conditions are not always associated to a combination of deltas. Recent works are directed towards detecting trend shifts in the variables [15]. Among them, some diagnostic methods are based on the detection of signatures that are combinations of slope changes in the EHM deltas known to be associated to specific events or conditions [8]. The distances between each of these signatures and a sequence of EHM values measured on an engine constitutes a feature vector that can be fed to a classifier in order to predict the deterioration level of an engine. However, it was found that some defects cannot be detected by a classifier operating under these principles, especially in the cases where the deterioration signatures are not yet known. This was resolved by using an all-inclusive catalogue of signatures, in combination with a sample of engines where all of the sought defects were present. Feature selection techniques were subsequently applied in order to identify the most relevant signatures, or alternatively a classifier could also be applied to implicitly perform the required feature selection [7]. In particular, the classifier in this last reference is a Fuzzy Rule-Based System (FRBS) whose Knowledge Base (KB) comprises rules of the following form:

```
IF TURBINE TEMPERATURE DECREASE
  AND FUEL FLOW INCREASE THEN
    COMPRESSOR HEALTH IS LOW
```

These techniques constitute an effective diagnosis system, able to detect the presence of abnormal events or significant engine conditions. However, the prediction of an engine's remaining life to a known condition (the prognosis problem previously mentioned) is a wider problem. An engine that repeatedly operates under unfavourable conditions has smooth levels of deterioration over time which inherently shorten the engine's life. Smooth deterioration trends do no manifest themselves as combinations of EHM signals, as a result the current existing techniques cannot be used to identify these deterioration trends.

In this paper a solution to this problem is presented which is based on a deterioration rate $r(t)$ model of a component as a function of the EHM variables. It is proposed that $r(t)$ is defined as the solution to the following integral equation:

$$\text{Remaining cycles}(t) = \text{Initial life} - \int_0^t r(\tau)\mathrm{d}\tau \quad (1)$$

For example, if the HPC has a constant deterioration rate $r(t) = 2$, and the initial life is of 5000 cycles, then the engine should undergo maintenance in 2500 cycles because $\text{Remaining cycles}(2500) = 0$. Deterioration rates lower than 1 are also considered, for those engines which are flying in above-average conditions. The cyclic or hourly remaing life calculation is dependent on the actual data available.

C. Distal learning of FRBS

Modelling the prognostic indicator through the integral of the instantaneous deterioration rate of an engine enables the identification of not only sudden events but also of smooth levels of deterioration, as previously mentioned. The simplest version of the estimator for the remaining cycles is obtained by assuming that the last known deterioration speed is constant throughout the remaining future cycles and solving Eq. 1 to determine the value T_0 for which $\text{Remaining cycles}(T_0) = 0$. This and other estimators are discussed in Section III-D.

An FRBS is used to link EHM data to deterioration rates. Learning the KB of an FRBS requires a training dataset with samples of the input and output variables. In this problem, this set would typically consist of a sample of engine measurements which would link the EHM variables to the specific known deterioration rates. However, deterioration rate is not an observable parameter and as such this sample dataset cannot be compiled. The KB must therefore be indirectly learnt from the available information, this is

1) The sequence of EHM variables considered are those measured in the time lapse between two shop visits.
2) The remaining life is based on the condition of each component at the end of the sequence, which is determined through the inspections carried out at the engine shop visit.
3) An estimation of the release life of each component at the beginning of the sequence can be made after the first shop visit.

This indirect learning task can be deemed to be a type of supervised learning problems known as "Distal Learning" [4]. In this kind of problems (see Figure 3), target values are available for the distal variables (the "outcomes") but

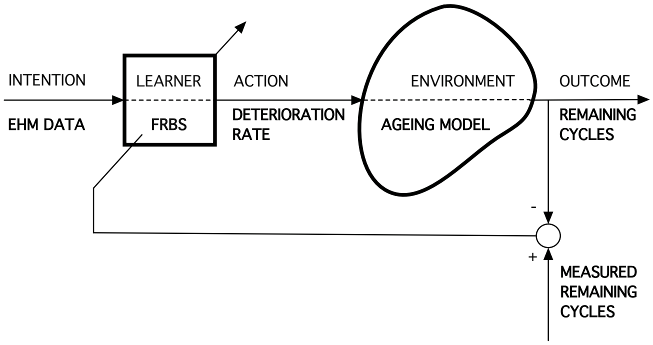

Fig. 3. Overview of the distal supervised learning problem. Target values are available for the distal variables (the "outcomes") but not for the proximal variables (the "actions") [4]. The target values are the life expectations measured at the shop visit. The proximal variables are the deterioration rates that are related to the distal variables through an ageing model. The ageing model has memory thus the outcome depends on the history of the actions.

not for the proximal variables (the "actions"). In the engine prognosis problem, the target values are the life expectations. The proximal variables are the deterioration rates, which are related to the distal variables through an ageing model of the engine. The ageing model has memory, thus the outcome depends on the history of the actions, i.e. the age of the engine depends on the sequence of deterioration rates. The learner, which in this case is the FRBS, previously mentioned, is adjusted so that the output of the ageing model at the end of an EHM data sequence matches the measured level of deterioration of the engine.

The proposed rule learning process is based on a Pitts Genetic Fuzzy System [2] where the fitness function is modified in order to include the ageing model. Distal learning has not been associated with Genetic Fuzzy Systems before, as far as we know, and as such additional details about the implementation of this specific combination are given in Section III.

The proposed KB comprises rules that map combinations of slope changes in EHM deltas and deterioration rates, in the following form:

```
IF TURBINE TEMPERATURE DECREASE
  AND FUEL FLOW INCREASE THEN
DETERIORATION RATE OF THE HPC IS LOW.
```

The main purpose the learnt FRBS is estimating the remaining cycles of the engine in combination with the ageing model mentioned. In this respect, the FRBS is a by-product of the learning task. However, in this particular application the FRBS is in itself a model of the instantaneous deterioration rate as a function of the EHM signals, which can in addition be used to gain an insight of the relationship between the values of the EHM variables and the engine's operating conditions. This will be discussed further in Section IV.

In short, this paper is structured as follows: the diagnosis problem is introduced in Section II. The proposed method is defined in Section III. Section IV contains a numerical analysis of the proposed algorithm against other alternatives. Section V concludes the work and discusses possible future research in the field.

II. EHM-BASED DIAGNOSIS OF AEROENGINES

A typical two shaft high bypass ratio turbo fan is depicted in Figure 1. In this type of engine, the thrust is performed by the air compressed by the fan blades and pushed through the engine bypass. The air pushed through the core of the engine is solely used to turn the fan. This is, the air is compressed by the high pressure compressor (HPC) so that the optimum conditions are reached within the combustion chamber to subsequently turn the high pressure turbine (HPT) to maintain the high pressure (HP) system and subsequently turn the low pressure turbine (LPT) which turns the fan and produces the engine thrust.

A. Stations in a turbofan

The main stations depicted in Figure 1 follow the most commonly used numbering convention. Although single digits are used to define the main stations, double digits are used to define interim positions. The first digit defines the main station whilst the second, defines an interim position.

- Station 2: Due to the design of the engine intake the temperatures and pressure at station 2 are different to those of station 0 and are more representative of the actual engine intake conditions which will be used as reference by the controls system. The main variables at this station are P2 and T2.
- Station 25: This is the entry to the HPC. Depending on the engine design a booster or an Intermediate Pressure Compressor (IPC) may also be associated to the low pressure (LP) system. As such station 25 is therefore defined as the entry to the HPC and not the exit of the fan.
- Station 3: This is the HPC exit and the entry into the combustion system. The conditions at this point are key for the correct functioning of the engine. The main variables measured at this station are P30 and T30.
- Station 4: This is the combustion chamber exit and HPT entry. The temperature at this point is one of the main engine parameters. T4, may also be known as Turbine Gas Temperature (TGT) or Internal Turbine Temperature (ITT)
- Station 5: This is the LPT exit. The main variable at this station is P50. This pressure is used to define EPR, which is subsequently used to determine the overall engine thrust. EPR is the relation of P50 to P20.

The LP system is the combination of the fan and the LPT. The speed at which the LP system turns is defined as N1. The HP system is the combination of the HPC and the HPT. The speed at which the HP system turns is known as N2. In addition, the amount of fuel consumed is also monitored through fuel flow (FF).

B. Engine deterioration

One of the main types of engine events or causes of deterioration is mechanical. Mechanical faults may be identified through overall engine deterioration and the assessment of EHM data. Independently of the system or component that is being assessed there are several stages or levels of deterioration throughout which the effect and associated costs and risk of continuous operation varies (recall the example in Figure 2). This is, any component or system will deteriorate over time solely due to its use, however if subject to an early inspection it could be identified to be good for further operation without maintenance. Further operation will deteriorate any component or system to a point at which if inspected would require the component or system to be repaired. Ultimately the level of deterioration of a component or system will reach a point where it will no longer be repairable. This unknown condition prior to the shop visit is in many cases is still safe for continuous operation. In many cases operational and maintenance costs will increase as the component or system is deteriorated and additional parts need to be replaced at the maintenance shop visit. In some cases, the system may deteriorate even further reaching an engine condition which could be deemed to be unreliable. In some cases material could also be released. In these cases high operational disruption and high maintenance costs would incurred as not only would the initial component be replaced but all of the secondary damaged components would also need to be repaired or replaced. In addition, the removal and maintenance of the engine would also need to be accommodated outside of their planned schedule. However the main issue in these situations is customer dissatisfaction and company reputation.

The main sections of any engine prone to significant events and deterioration are the high pressure compressor and turbine. This is where the air is compressed to the exact pressures required so that the fuel combustion can be optimized for improved efficiency and reduced pollution, with the turbine generating the work to keep the system running. As a consequence of this, these two engine systems or modules are the areas where the main maintenance costs are incurred. High Pressure Compressor or HPC deterioration is mainly driven by increased tip clearances, which in turn reduce the working line of the system, or by actual material release of a blade or a vane. Increased tip clearances may be induced by liner loss or by reduced blade height, either way increased clearances are a sign of deterioration [3]. High Pressure Turbine and Combustor deterioration may be due to the actual combustor been deteriorated, the fuel burn not been appropriate or actual blade or vane damage. Combustor deterioration is mainly time driven and is not typically identified through EHM methods due to its slow rate of deterioration. Turbine blade deterioration is mainly driven by reduced cooling or actual aerofoil cracking [11] which is either seen as an efficiency improved turbine or not actually visible through EHM signatures.

As a result, deterioration of HPC and HPT modules is expected to influence EHM data. A prognostic indicator of HPC and HPT remaining life through an EHM data assessment is therefore proposed in this paper. The main purpose of this indicator proposed is to determine the number of remaining flight cycles for the compressor and turbine modules respectively up to an agreed module condition which optimizes both engine time on-wing and maintenance costs.. The EHM subset of parameters considered in this study consists of the following five variables:

1) FF: Fuel flow
2) N2: Speed of the high pressure system
3) P30: High pressure compressor exit pressure
4) T30: High pressure compressor exit temperature
5) TGT: Turbine gas temperature

III. PROPOSED METHOD

An algorithm which is used to learn the expression of a prognostic indicator using Genetic Fuzzy Systems (GFSs) is proposed in this section. The training data consists of historical EHM data from sampled engines from the same fleet but from different operators and regions ie. from different flight conditions.

The method proposal is exposed four parts, detailing

- the procedures for cleaning, discretizing and transforming the uncertain input data into a sequence of fuzzy numbers
- the structure of the FRBS that is learnt
- the fitness function that the Genetic Algorithm (GA) is required to optimize including the definition of the ageing model
- the definition of the prognostic indicator in terms of the learnt FRBS.

An overview schematic of the process is shown in Figure 4.

A. Cleaning, discretizing and transforming input data

EHM data is very noisy and is not expressed in absolute values. The state of an engine is estimated from the deltas between an engine's own measurements and those from a known baseline engine, as previously discussed. It will be assumed that the deterioration rate depends on the speed of change of the EHM signals, as such the deterioration rate model will in turn be fed with the derivative of these signals.

Estimating the derivative of a noisy signal requires the use of low-pass filters that remove the high frequency content. In particular, it is proposed that the derivatives of the EHM signals are approximated by locally fitting straight lines to the smoothed EHM data. The smoothing will in turn be carried out with a kernel filter. For instance, let the temperature of the turbine TGT be the signal considered for this assessment. The smoothed value of this signal is given by the convolution of TGT with a Gaussian kernel function K, whose bandwidth Δ is related to the cut-off frequency of the filter:

$$\widehat{\mathrm{TGT}}(t) = \sum_{\tau=-\tau_0}^{\tau_0} \mathrm{TGT}(t+\tau) \cdot K(\tau, \Delta). \qquad (2)$$

Estimating the derivative of TGT is carried through the slope of a line locally fitted to $\widehat{\mathrm{TGT}}$. This line can be determined by weighted least squares. Given the values of time t and

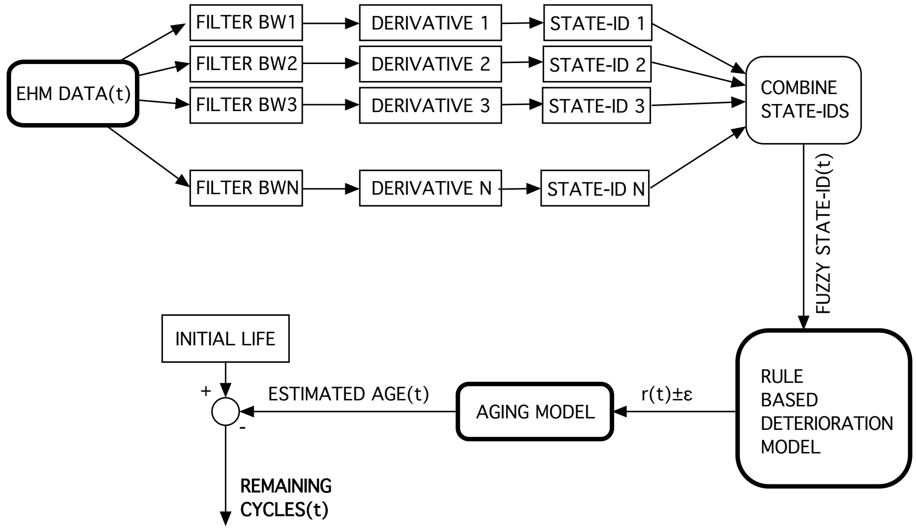

Fig. 4. Block diagram of the proposed method for estimating the remaining cycles of an aeroengine with EHM data

bandwidth Δ, the slope a and the y-intercept b of the best-fit line are at the minimum of the following function:

$$err(a, b) = \sum_{\tau=-\tau_0}^{\tau_0} \widehat{\mathrm{TGT}}(t+\tau) - (a\tau + b))^2 \cdot K(\tau, \Delta). \quad (3)$$

The sequence of slopes a(t) is therefore an estimate of the derivative dTGT/dt in this particular example, or the derivative of an arbitrary health. In this paper five derivatives are considered through this means: dTGT/dt, dFF/dt, dP30/dt, dT30/dt, and dN2/dt. In the following, the values of these five derivatives will be referred to as the state of the engine.

Since a rule-based model is to be used, the state must be discretized and a finite set of combinations defined. Each numerical value of a derivative will therefore be replaced by a label. The linguistic labels defined will be either “DOWN”, “SAME” or “UP”. A soft discretization is performed: if the state is x_0, and L is a linguistic label, the degree of truth of the assert “x_0 is L” is understood as a possibility $\Pi_L(x_0) = \mu_L(x_0)$. Observe that this possibilistic setup is also valid for the uncertain EHM signal measurements; the degree of truth of the assert “$x_0 \pm \epsilon$ is L” is $\Pi_L(x_0 \pm \epsilon) = \sup_{x\in[x_0-\epsilon,x_0+\epsilon]} \mu_L(x)$. For instance, the following is a valid discrete value of TGT:

$$\mathtt{TGT} = \{\mathtt{UP/0.8,SAME/0.3}\} \quad (4)$$

As a corollary of this kind of uncertainty representation, missing values have membership 1 to all labels.

Each set of 5 linguistic labels will be assigned a number. This number will be called the “State-Id”. In this case, with three possible slopes and considering the five variables above, there are 243 different possible State-Ids (three to the power of five). A base-3 numbering scheme is used, where the digits down=0, same=1, up=2 are respectively assigned to each label. For instance, the set of labels (down, same, up, up, down) would be assigned in base-3 the number 01220, whose corresponding State-Id is 51 in base 10.

Observe that each combination of EHM variables is not assigned a precise State-Id but a fuzzy subset of all the possible Ids as a result of the soft discretization. In turn, this subset is also dependent upon the selected bandwidth. In this respect, it was decided not to choose an arbitrary value for the bandwidth but to sweep a range of bandwidths and combine their corresponding fuzzy State-Ids into a discrete sequence that is to be subsequently fed to the deterioration rate model.

The numerical procedure for sweeping the range of bandwidths is based on a Monte-Carlo simulation with multiple repetitions of the whole filtering and discretization process, for different values of Δ. The set of values obtained are combined into a single fuzzy set, whose membership defines a possibility distribution over the set of State-Ids, following the procedure defined in [8]. After this, the EHM data of an engine is reduced to a chain of fuzzy numbers

$$\widetilde{\mathrm{StateId}}(t) = (\mu_1(t), \mu_2(t), \ldots, \mu_{243}(t)) \quad (5)$$

This chain is the input to the rule-based model used to predict the specific HPC and HPT deterioration rate.

B. Structure of the FRBS modelling the deterioration rate

Two different FRBSs have to be learnt, to model the HPC and HPT respectively. Each of them has five inputs, dTGT/dt, dFF/dt, dP30/dt, dT30/dt, and dN2/dt. As discussed before, each input is discretized into the linguistic labels “down”, “same” and ‘up”. Mamdani-type rules are used, for instance:

```
IF dTGT/dt=SAME AND dFF/dt=UP AND dP30/dt=UP
AND dT30/dt=DOWN AND dN2/dt=UP THEN
DETERIORATION RATE OF THE HPC IS LOW
WITH CONFIDENCE FACTOR 0.8
```

which is the same as

```
IF STATE-ID=12202(3 THEN
DETERIORATION RATE OF THE HPC IS LOW
WITH CONFIDENCE FACTOR 0.8
```

Observe that neither fuzzyfication nor defuzzification interfaces are needed in the proposed system. The degree of truth of the k-th antecedent is the membership value $\mu_k(t)$ in the input chain of fuzzy numbers $\widetilde{\mathrm{StateId}}(t)$ previously described.

The output of each FRBS is not a number but an interval $\overline{r}(t) = [r^-(t), r^+(t)]$ because the input is not crisp. Given that the fuzzy State-Id was given a possibilistic interpretation, this output interval ranges the possible outputs of the FRBS when the degrees of truth of the rules in the KB are the probability distributions dominated by the possibility distribution of State-Ids,

$$\overline{r}(t) = \left\{ \sum_{k=1}^{243} p_k \cdot \omega_k \cdot R_k \mid \sum_{k=1}^{243} p_k = 1,\ 0 \le p_k \le \mu_k(t) \right\} \tag{6}$$

where R_k and ω_k are the modal point of the linguistic label in the k-th consequent and the weight of the rule whose antececent refers to the k-th State-Id, respectively. This interval of values is passed on to the ageing model in order to compute the fitness function.

C. Ageing model and fitness function

The most simple form of the ageing model consists in integrating the deterioration rate over time. The number of remaining cycles is

$$\text{Cycles}(t) = \text{Initial Life} - \text{Estimated Age}(t) \tag{7}$$

Given that $\overline{r}(t) \subset [0, \infty)$, the following holds:

$$\int_0^{t_0} r^-(\tau)\, \mathrm{d}\tau \le \text{Estimated Age}(t) \le \int_0^{t_0} r^+(\tau)\, \mathrm{d}\tau \tag{8}$$

In practical cases, the ageing model must also take into account engine events (which may cause a sudden change to the estimated age) or even an on-wing maintenance operation. The discrete form of the ageing model is therefore

$$\begin{aligned} \text{Cycles}(k) = {} & \text{Initial Life} + \\ & + \sum_{\tau=0}^{k} (\text{maintenance}(\tau) - \text{events}(\tau)) \\ & - \frac{1}{2} \sum_{\tau=0}^{k} (r^+(\tau) + r^-(\tau)) \\ & \pm \frac{1}{2} \sum_{\tau=0}^{k} (r^+(\tau) - r^-(\tau)) \end{aligned} \tag{9}$$

Therefore, given a sample of N aeroengines whose expected life was f_i when inspected after c_i cycles, the fitness of the FRBS may be evaluated by means of an interval-valued function, as follows:

$$\text{fit} = \left\{ \sum_{i=1}^{N} |t_i - f_i| : t_i \in \text{Cycles}(c_i) \right\} \tag{10}$$

With respect to the encoding mechanism in the GA, and given that each of the KBs is made up by a maximum of 243 rules, all parameters can be jointly encoded in the same genotype (Pitts-style GFS) with a reasonable computational efficiency. However, it is remarked that a nonstandard GA is required in order to optimize Eq. 10 and determine the parameters which define the KB. This is because the proposed fitness function is not numerical but interval-valued. The algorithm proposed in [12], [13] was used. Lastly, observe that it was decided not to tune the membership functions of the labels "UP", "SAME" and "DOWN" but to weight the fuzzy rules instead.

D. Definition of the prognostic indicator

The prognosis indicator is intended to estimate the remaining life of an engine, through a prediction of its deterioration rate. Extrapolating these rates is deemed will allow to dynamically re-schedule the maintenance checks of engines with higher and lower than normal deterioration rates, anticipating certain events or costly findings thus reducing the number or degree of unforeseen engine shop visits.

For an extrapolated rate $\widehat{r}(\tau)$ for $\tau > t$, it is proposed that the prediction at time t of the useful life $T(t)$ of an engine is the solution to the following integral equation:

$$\text{Initial life} - \int_0^t r(\tau)\mathrm{d}\tau - \int_t^T \widehat{r}(\tau)\mathrm{d}\tau = 0 \tag{11}$$

In this work a 0-th order prognosis indicator $T_0(t)$ was used. This considers a constant rate of deterioration rate $\widehat{r}(\tau) = r_0$ for $\tau > t$, thus

$$T_0(t) = t + \frac{\text{Initial life} - \int_0^t r(\tau)\mathrm{d}\tau}{r_0}. \tag{12}$$

Different strategies can be used for assigning a value to r_0: the last known rate $r(t)$, the average deterioration $r_0 = 1/t \cdot \int_0^y r(t)\mathrm{d}t$ or the unity value, to name a few. Higher order prognosis models can be defined by using time series models to extrapolate $r(t)$ or the EHM variables, however it was found that the accuracy of the higher order models does not significantly improve the 0-th order model with extrapolated unity deterioration rate.

IV. Numerical results and discussion

The level of deterioration of an engine is determined through the inspections carried out at the engine shop visit. The cycles at which certain events or findings occur are not all known, thus a training sample made up of only of engines with smooth levels of deterioration is not possible. As a consequence of this, a training dataset comprising 43 engines without a detectable signature was compiled. The experimental design in this section is therefore guided to compare the results of a state-of-the-art signature-based regression model against the proposed approach. It will be shown that the regression model is not better in this sample than a purely periodical maintenance schedule, but there is a statistically significant difference favouring Genetic Distal Learning. This result will be used to assess those types of deterioration which are not detectable though existing EHM signatures as well as establish that the proposed algorithm can successfully diagnose most levels of deteriorations.

A. Compared results

The procedure described in [7], except for the classification stage, has been applied first to the sample of 43 engines as previously described. The aforementioned classification stage was replaced by a regression module that approximates the expected life of either the HPC or the HPT. The input variables are the same feature vector used in the removed classification stage. Random forests were used for the regression task [1]. It is also highlighted that the results from the analysis of the dispersion of the classification which was studied in this last reference, have not been carried to the regression model defined, and only make use of the centroids of the mentioned feature vector.

As a reference of the quality of the prognosis models, a naive model has also been considered where the deterioration rate was determined to be constant and equal to 1. In other words, the expected life of the engine is considered as the difference between the initial life of the module and the number of cycles the engine has flown. It is remarked that this is the standard procedure based on average service experience typically used to schedule maintenance checks.

Lastly, the Genetic Distal Learning of a FRBS was combined with a 0-th order prognosis indicator and a unity extrapolated deterioration rate. A 10-cv validation was used in all comparisons. The compared results are shown in Table I. Observe that Distal Learning is the best alternative for both HPC and HPT, however the accuracy gain of the method with respect to the standard scheduling is better for compressors (20% on average) than for turbines (4%).

TABLE I. AVERAGE ACCURACY (10-CV) FOR HPC AND HPT USING A DISTAL LEARNING, A SIGNATURE-BASED RANDOM FOREST REGRESSION MODEL AND THE STANDARD PROCEDURE

Method	HPC	HPT
Distal	**1330**	**1541**
Signature	1426	1558
Standard	1651	1579

The relevance of the differences between the methods are illustrated in Figures 5, 6 and 7. Figure 5 shows three boxplots with the dispersion of the 10-cv test results with the absolute differences between the HPC predicted life and the measured values for Distal, Signature-based and Standard techniques in HPC. The same boxplots are shown for the HPT in Figure 6.

The p-values of the paired differences between the standard method and the proposed algorithm are negligible for both HPC and HPT, although the percent gain is much higher for compressors, as mentioned. A boxplot with these paired differences is shown in Figure 7. This figure serves also as a justification of the p-value found in the statistical tests about the difference of the mean accuracy of both algorithms; observe that all differences are lower or equal than zero, meaning that Distal Learning improved the standard scheduled maintenance for all folds in the validation test.

In addition, in Figure 8 the unfiltered EHM signals are shown, along with their filtered derivatives for a particular bandwidth, as well as the the outputs of the deterioration rate models and the outputs of the prognostic indicators. The

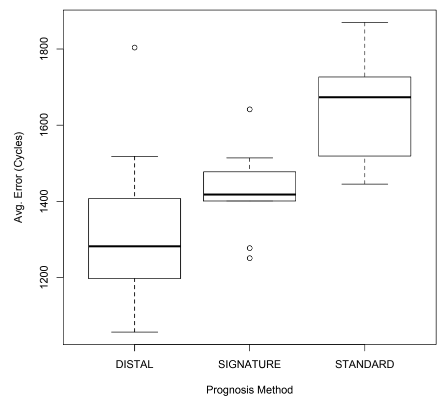

Fig. 5. Dispersion of the 10-cv test results with the absolute differences between the predicted life and the measured values for Distal, Signature-based and Standard techniques in HPC

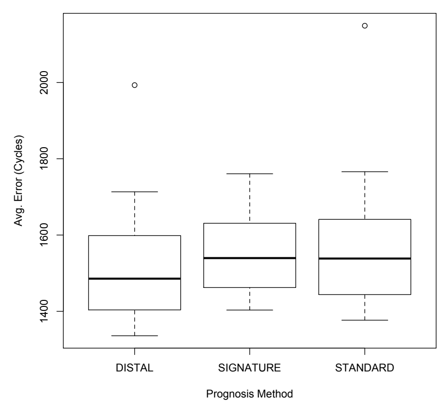

Fig. 6. Dispersion of the 10-cv test results with the absolute differences between the predicted life and the measured values for Distal, Signature-based and Standard techniques in HPT

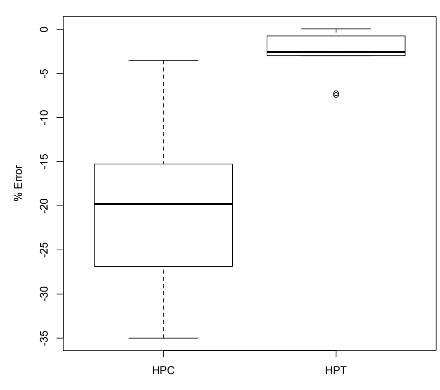

Fig. 7. Boxplot of the paired differences between Standard and Distal algorithms, showing that the proposed algorithm improved the standard maintenance schedule for all folds in the validation.

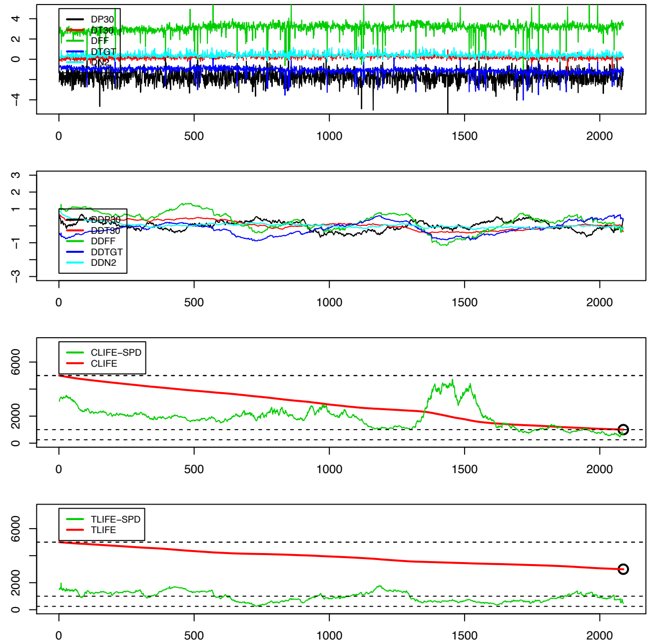

Fig. 8. From upper to lower: EHM signals, slopes of the filtered EHM signals for a given bandwidth, HPC and HPT deterioration rates and prognostic indicators. Green curves in the last two plots are the deterioration rates, red curves are the expected life of the components.

green curves in the two plots in the lower part of the figure are the outputs of the deterioration rate model. Observe that the combination of EHM signals around sample 1500 show a particularly harsh set of conditions for the compressor, and that generally speaking the fast deteriorations of compressor and turbine alternate in time. The red curves are the integral of the deterioration, assuming that the initial life of HPC and HPT was 5000 cycles. The circles at the end of the red curves are the measured life of these elements as observed at the shop visit. The difference between the height of these circles and the red curves are the centerpoint of the fitness function defined in the preceding section.

V. Concluding Remarks

This work shows potential to predict the remaining life of an engine through the use of EHM data applying Genetic Distal Learning techniques. Generally speaking, most of the engines can be diagnosed with existing techniques, but there are certain types of defects that do not manifest themselves as a change in the slope of the EHM data but as a smooth deterioration that cannot be detected.

The supervised learning with a distal teacher paradigm, adapted for uncertain data and genetic algorithms, has been used to learn FRBS from sequences composed of fuzzy discretizations of the different EHM variables. These FRBS are used to predict the deterioration rate of HPC or HPT in an aeroengine. An ageing model that integrates these instantaneous deteriorations is devised which produces an online estimation of the remaining life of the engine. As a by-product of the learning process, the FRBS shows the combinations of EHM values that are associated with an increased level of deterioration for HPC or HPT and therefore detects the cycles where the deterioration was higher. The opposite is also true for those cases where reduced level of deterioration are incurred. The results have been tested with a representative sample of planes. It was determined that the results of previous prognostic methods can be improved by including the new algorithm in the existing available catalogue of assessment techniques.

Acknowledgements

This work was supported by the Spanish Ministerio de Economía y Competitividad under Project TIN2011-24302, including funding from the European Regional Development Fund.

References

[1] L. Breiman, Random Forests, Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.

[2] O. Cordon, F. Herrera, F. Gomide, F. Hoffmann, and L. Magdalena, Ten years of genetic fuzzy systems: current framework and new trends, presented at the IFSA World Congress and 20th NAFIPS International Conference, 2001. Joint 9th, 2001, vol. 3, pp. 1241-1246.

[3] M. Foerstemann and S. Staudacher, Optimizing the Architecture of Civil Turbofan Engines to Improve Life Cycle Costs/Value Added, ASME Turbo Expo 2004: Power for Land, Sea, and Air, pp. 89-96.

[4] M. I. Jordan and D. E. Rumelhart, Forward models: Supervised learning with a distal teacher, Cognitive Science, 1992.

[5] A. Kyriazis, A. Tsalavoutas, K. Mathioudakis, M. Bauer, and O. Johanssen, Gas Turbine Fault Identification by Fusing Vibration Trending and Gas Path Analysis, ASME Turbo Expo 2009: Power for Land, Sea, and Air, pp. 687-696.

[6] Y. G. Li, Gas Turbine Performance and Health Status Estimation Using Adaptive Gas Path Analysis, ASME Turbo Expo 2009: Power for Land, Sea and Air, 2009, pp. 1-13.

[7] A. Martínez, L. Sanchez, and I. Couso, Engine Health Monitoring for engine fleets using fuzzy radviz, 2013 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 1-8.

[8] A. Martínez, L. Sánchez, I. Couso. Interval-valued Blind Source Separation applied to AI-based prognostic fault detection of aircraft engines. Journal of Multiple Valued Logic and Soft Computing. In press.

[9] M. Müller, S. Staudacher, W. H. Friedl, R. Köhler, and M. Weißschuh, Probabilistic Engine Maintenance Modeling for Varying Environmental and Operating Conditions, ASME Turbo Expo 2010: Power for Land, Sea, and Air. ASME, 2010.

[10] M. Naeem, Implications of engine deterioration for a high-pressure turbine-blade's low-cycle fatigue (LCF) life-consumption, International Journal of Fatigue, vol. 21, no. 8, pp. 831-847, Sep. 1999.

[11] S. O. T. Ogaji, S. Sampath, R. Singh, and S. D. Probert, Parameter selection for diagnosing a gas-turbine's performance-deterioration, Applied Energy, vol. 73, no. 1, pp. 25-46, Sep. 2002.

[12] L. Sanchez, I. Couso, J. Casillas, Genetic learning of fuzzy rules on low quality data, Fuzzy Sets and Systems vol. 160, no. 17, pp. 2524-2552, 2009.

[13] L. Sánchez and I. Couso, Advocating the use of imprecisely observed data in genetic fuzzy systems, IEEE Trans. Fuzzy Syst., vol. 15, no. 4, pp. 551-562, 2007.

[14] S. Spieler, S. Staudacher, R. Fiola, P. Sahm, and M. Weißschuh, Probabilistic Engine Performance Scatter and Deterioration Modeling, ASME Turbo Expo 2007: Power for Land, Sea, and Air. ASME, pp. 1073-1082.

[15] L. A. Urban, Gas Path Analysis Applied to Turbine Engine Condition Monitoring, Journal of Aircraft, vol. 10, no. 7, 2012.